Action-Conditioned 3D Human Motion Synthesis with Transformer VAE

Year: 2021 | Citations: 604 | Authors: Mathis Petrovich, Michael J. Black, Gul Varol

Abstract

We tackle the problem of action-conditioned generation of realistic and diverse human motion sequences. In contrast to methods that complete, or extend, motion sequences, this task does not require an initial pose or sequence. Here we learn an action-aware latent representation for human motions by training a generative variational autoencoder (VAE). By sampling from this latent space and querying a certain duration through a series of positional encodings, we synthesize variable-length motion sequences conditioned on a categorical action. Specifically, we design a Transformer-based architecture, ACTOR, for encoding and decoding a sequence of parametric SMPL human body models estimated from action recognition datasets. We evaluate our approach on the NTU RGB+D, HumanAct12 and UESTC datasets and show improvements over the state of the art. Furthermore, we present two use cases: improving action recognition through adding our synthesized data to training, and motion denoising. Code and models are available on our project page [53].